

17/F

FILED
IN CLERK'S OFFICE
U.S. DISTRICT COURT E.D.N.Y.
★ NOV 26 2019 ★
BROOKLYN OFFICE

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

-----X

UNITED STATES OF AMERICA,

-against-

ALONZO SHIPP,

Defendant.

-----X

NICHOLAS G. GARAUFIS, United States District Judge.

MEMORANDUM & ORDER

19-CR-029 (NGG)

Defendant Alonzo Shipp moves to exclude testimony from the Government's proposed ballistics expert, Detective Sean Ring. (See Mot. to Exclude (Dkt. 33); Mem. in Supp. of Mot. to Exclude ("Mem.") (Dkt. 33-2).) In the alternative, Mr. Shipp requests that the court limit the expert testimony. (Mem. at 15-16). For the following reasons, the court DENIES Defendant's motion to exclude, but GRANTS his request to limit Detective Ring's testimony.

The parties agree that no court has entirely excluded expert testimony on firearms toolmark analysis, although courts frequently do place limitations on the level of certainty the expert may profess. To overcome this case law, Mr. Shipp relies primarily on a recent report from the President's Council of Advisors on Science and Technology ("PCAST"), which reviewed the available research on firearms toolmark analysis and found that the method lacks foundational scientific validity.

The court has carefully considered the PCAST Report, the earlier National Research Council ("NRC") Report, other applicable scientific literature, and relevant case law, and has determined that Detective Ring may testify as an expert in the field of firearms toolmark analysis. However, because the PCAST Report's findings cast considerable doubt on the reliability of the theory behind matching pieces of ballistics evidence, Detective Ring will be

permitted to testify only that the toolmarks on the recovered bullet fragment and shell casing are consistent with having been fired from the recovered firearm. In other words, Detective Ring may testify that the recovered firearm cannot be excluded as the source of the recovered bullet fragment and shell casing, but not that the recovered firearm is, in fact, the source of the recovered fragment and shell casing. Additionally, Detective Ring may testify based on his knowledge, training, and experience about his method for analyzing and test firing the recovered firearm, the procedure for comparing the test fires to the recovered bullet fragment and shell casing, and the similarities he observed between the recovered ballistics evidence and the test fires from the recovered firearm.

I. BACKGROUND

A. Facts¹

1. Alleged Shooting

On or about July 20, 2018, an unnamed individual, referred to herein as John Doe, was shot in the vicinity of 117-26 147th Street in Queens, New York. (Compl. (Dkt. 1) ¶ 2.) Doe then ran south down 147th Street and east on 119th Avenue to the corner of 119th Avenue and Sutphin Boulevard, where he collapsed. (Id.) The NYPD later recovered a shell casing around 117-26 147th Street. (Id.)

Video footage then shows the gunman approaching Doe, taking an item out of his pants or waistband, standing over Doe, pointing an object at him, and then walking away. (Id. ¶¶ 5, 7.) Video recovered from a surveillance camera at the corner of Sutphin Boulevard and Foch Boulevard (roughly two blocks north of where Doe was found) shows an individual wearing

¹ The court assumes the parties' familiarity with the factual background and sets forth here only those facts relevant to Mr. Shipp's motion to exclude Detective Ring's testimony.

similar clothes as the gunman walking north on Sutphin Boulevard and pausing outside a business located at the corner. (Id. ¶ 9.) The next morning, an employee of that business discovered a 9mm Sig Sauer handgun in the dumpster in front of the business. (Id. ¶ 10.) The firearm had “ten rounds in the magazine and one spent shell casing jammed in the ejection port.” (Id.)

On January 2, 2019, Mr. Shipp was arrested and charged with possession of the firearm alleged to have been used in the July 20, 2018 incident. (See Indictment (Dkt. 7) ¶ 1.)

2. The Ballistics Evidence

After the shooting, NYPD personnel collected and processed physical evidence from the crime scene, including bullet fragments and shell casings. (Gov’t Mem. in Opp’n to Mot. to Exclude (“Gov’t Opp’n”) (Dkt. 34) at 3); (see also Decl. of Ashley M. Burrell in Supp. of Mot. to Exclude (Dkt. 33-1) ¶ 3). This evidence was analyzed by the Firearms Analysis Section of the NYPD Police Laboratory, including Detective Ring. (Gov’t Opp’n at 3.)

Detective Ring initially analyzed two bullet fragments and one cartridge casing recovered from the crime scene. (See Shipp Discovery (Dkt. 33-5) Bates No. ASHIPP000293 (“Shipp 293”).) His initial analysis determined that the casing and one fragment were suitable for microscopic comparison, while the second fragment lacked “discernible class and/or individual characteristics.” (Id.) Ring’s notes do not indicate the provenance of these pieces of evidence. (Id. at 295-96.) Ring later analyzed four additional bullet fragments that were recovered from Doe’s body. (Id. at 300.) He found them to be suitable for comparison, but, for reasons not apparent from the record, he apparently did not perform a comparison on these fragments. (Id. at 298-301.)

Detective Ring also test fired the recovered firearm and analyzed one casing and four bullets. (Id. at 289-92.) Ring then compared the test fires to the casing and bullet fragment recovered from the crime scene. (Id. at 279-84.) He concluded that the cartridge casing recovered at the crime scene was fired from the recovered firearm “based on the observed agreement of their class characteristics and sufficient agreement of their individual characteristics.” (Id. at 279.) He concluded the bullet fragment was fired from the recovered firearm for identical reasons. (Id.) The documents provided contain no additional information explaining which marks Ring relied on to conclude there was “sufficient agreement” of the individual characteristics between the test fires and the recovered ballistics evidence. (See id. at 279-284.)

B. Procedural History

On January 2, 2019, Mr. Shipp was arrested and charged with possession of the firearm alleged to have been used in the July 20, 2018 incident. (See Indictment (Dkt. 7) ¶ 1.) He was denied bail on January 10, 2019. (See Jan. 10, 2019 Min. Entry (Dkt. 4); Order of Detention (Dkt. 5).) Mr. Shipp was arraigned on February 1, 2019 before Magistrate Judge Scanlon, at which point he entered a plea of not guilty. (Feb. 1, 2019 Min. Entry (Dkt. 11).)

By letter on August 1, 2019, the Government disclosed its intent to call Detective Ring as a ballistics expert. (Aug. 1, 2019 Gov’t Letter (Dkt. 32) at 1-2.) On August 16, 2019, Defendant moved to exclude Detective Ring’s testimony as unreliable under Federal Rule of Evidence 702 and Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), and requested a Daubert hearing to determine the reliability of Detective Ring’s testimony. (See Mot.; Mem.) In the alternative, Defendant moved to limit Detective Ring’s testimony. (Mem. at 15-16.) The Government opposed the motion on September 13, 2019 (Gov’t Opp’n.), and

Defendant entered a reply on September 26, 2019 (Reply in Mem. in Support of Mot. to Exclude (“Reply”) (Dkt. 35)).

The court heard oral argument on the motion on October 3, 2019 (see Oct. 3, 2019 Min. Entry) and ordered the government to provide any additional notes, written summaries or sketches related to Detective Ring’s conclusion in this case, as well as information about Detective Ring’s proficiency testing as a firearms toolmark examiner. (Id.) Pursuant to this order, the Government filed a supplemental response on October 11, 2019 (Suppl. Submission in Resp. to Court Order (Dkt. 36)), and Defendant replied on October 18, 2019 (Reply to Gov’t Suppl. Submission (Dkt. 37)).

II. LEGAL STANDARD

The admissibility of Detective Ring’s testimony is governed by Federal Rule of Evidence 702. Under the Rule, a witness may testify as an expert if they are qualified “by knowledge, skill, experience, training, or education.” Fed. R. Evid. 702. If a witness qualifies as an expert, Rule 702 allows for their testimony when:

- a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- b) the testimony is based on sufficient facts or data;
- c) the testimony is the product of reliable principles and methods;
and
- d) the expert has reliably applied the principles and methods to the facts of the case.

Id. Daubert clarified that the district court holds the gatekeeping function of “ensuring that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand.” 509 U.S. at 597. The court’s gatekeeping role extends “not only to testimony based on ‘scientific’

knowledge, but also to testimony based on ‘technical’ and ‘other specialized’ knowledge.”

Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 141 (1999) (quoting Fed. R. Evid. 702).

To gauge the reliability of proffered testimony, “the district court should consider the indicia of reliability identified in Rule 702,” which are not exhaustive. Wills v. Amerada Hess Corp., 379 F.3d 32, 48 (2d Cir. 2004). In doing so, “the district court has broad discretion in determining what method is appropriate for evaluating reliability under the circumstances of each case.” Restivo v. Hessemann, 846 F.3d 547, 575 (2d Cir. 2017) (quoting Amorgianos v. Nat’l R.R. Passenger Corp., 303 F.3d 256, 265 (2d Cir. 2002), cert. denied, 138 S. Ct. 644 (2018)). Courts routinely consider the five additional factors listed in Daubert as a starting point. These are: (1) “whether [the] theory or technique . . . can be (and has been) tested;” (2) “whether the theory or technique has been subjected to peer review or publication;” (3) “in the case of a particular scientific technique, the known or potential rate of error,” (4) “the existence and maintenance of standards controlling the technique’s operation;” and (5) whether a particular technique or theory has gained “general acceptance.” Daubert, 509 U.S. at 593-94. The district court’s inquiry is “flexible,” id at 594; the Daubert factors “neither necessarily nor exclusively appl[y] to all experts or in every case,” Kumho Tire, 526 U.S. at 141.

“[T]o ‘warrant admissibility[,] it is critical that an expert’s analysis be reliable at every step.’” United States v. Morgan, 675 F. App’x 53, 55 (2d Cir. 2017) (summary order) (quoting Amorgianos, 303 F.3d at 267). Expert testimony may be excluded if “there is simply too great an analytical gap between the data and the opinion proffered.” Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997). “Frequently, though, ‘gaps or inconsistencies in the reasoning leading to the expert’s opinion go to the weight of the evidence, not to its admissibility,” Restivo, 846 F.3d at 577 (quoting Campbell ex rel. Campbell v. Metro. Prop. & Cas. Ins. Co., 239 F.3d 179, 186 (2d

Cir. 2001) (alterations adopted)), and courts should remember that “vigorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.” Morgan, 675 F. App’x at 55 (quoting Daubert, 509 U.S. at 596). However, “the explicit premise of Daubert and Kumho Tire is that, when it comes to expert testimony, cross-examination is inherently handicapped by the jury’s own lack of background knowledge.” United States v. Ashburn, 88 F. Supp. 3d 239, 248-49 (E.D.N.Y. 2015) (Garaufis, J.) (quoting United States v. Glynn, 578 F. Supp. 567, 574 (S.D.N.Y. 2008)).

Finally, the party proffering expert testimony bears the burden of demonstrating, by a preponderance of the evidence, that the expert testimony is admissible. United States v. Williams, 506 F.3d 151, 160 (2d Cir. 2007); Fed. R. Evid. 702 Advisory Committee’s note to the 2000 Amendment.

III. DISCUSSION

First, the court will discuss Defendant’s request for a Daubert hearing on the admissibility of Detective Ring’s testimony. Then, it will review Detective Ring’s expert qualifications. Next, it will explain the theory of toolmark analysis and discuss two recent scientific reports that have analyzed the method’s reliability. Finally, the court will apply the Daubert factors to toolmark analysis.

A. Defendant’s Request for a Daubert Hearing

A district court is not required to hold a formal Daubert hearing in advance of qualifying an expert witness. See Williams, 506 F.3d at 161 (“While the gatekeeping function requires the district court to ascertain the reliability of [the expert’s] methodology, it does not necessarily require that a separate hearing be held in order to do so.”); Ashburn, 88 F. Supp. 3d at 244 (“Nothing requires a district court to hold a formal Daubert hearing in advance of qualifying an

expert witness.”). In addition to the parties’ papers and oral argument, the court has reviewed prior judicial decisions concerning the reliability of toolmark analysis, the PCAST and NRC reports, several of the individual studies discussed in these reports, and relevant academic articles discussing toolmark analysis. These materials provide the court with a thorough and well-documented record regarding the proposed testimony, firearms toolmark analysis, and the AFTE Theory of Identification. Accordingly, Defendant’s request for a separate Daubert hearing is DENIED.

B. Detective Ring’s Qualifications

Whether a witness qualifies as an expert under Rule 702 is a “threshold question” that should be considered before the “reliability inquir[y].” Nimely v. City of New York, 414 F.3d 381, 396 n.11 (2d Cir. 2005). A court should consider the purported expert’s “background and practical experience” to determine whether he is qualified to testify as an expert. See McCulloch v. H.B. Fuller Co., 61 F.3d 1038, 1043 (2d Cir. 1995).

Detective Ring joined the NYPD in 1999 and has been a Detective 3rd Grade since 2012. (Curriculum Vitae of Detective Sean Ring (“Ring CV”) (Dkt. 33-4).) He is assigned to the Firearms Analysis section of the NYC Police Laboratory. (Id.) In the course of his work, he has “analyzed and tested the operability of over 1500 firearms” and “microscopically examined thousands of pieces of ballistics evidence.” (Id.) He has completed multiple trainings in the area of toolmark analysis, including the two-and-a-half-year NYPD Firearms Examiner Training Program, and he now trains others in the proper procedures around conducting ballistics examinations. (Id.) Finally, Detective Ring has previously been qualified to testify as an expert in the Southern and Eastern Districts of New York. (Id.)

Detective Ring is therefore qualified to testify as an expert based on his “background and practical experience.” McCulloch, 61 F.3d at 1043; see also Fed. R. Evid. 702.

C. Firearms Toolmark Analysis

1. Firearms Toolmark Analysis Theory and Background

United States v. Otero, 849 F. Supp. 2d 425 (D.N.J. 2012), aff'd, 557 F. App'x 146 (3d Cir. 2014) described the underlying theory of toolmark analysis as follows:

Toolmark identification is based on the theory that tools used in the manufacture of a firearm leave distinct marks on various firearm components, such as the barrel, breech face or firing pin. The theory further posits that the marks are individualized to a particular firearm through changes the tool undergoes each time it cuts and scrapes metal to create an item in the production of the weapon. Toolmark identification thus rests on the premise that any two manufactured products, even those produced consecutively off the same production line, will bear microscopically different marks. With regard to firearms, these toolmarks are transferred to the surface of a bullet or shell casing in the process of firearm discharge. Depending on the tool and the type of impact it makes on the bullet or casing, these surface marks consist of either contour scratch lines, known as striations (or striae), or impressions. For example, rifling (spiraled indentations) inside of a gun barrel will leave raised and depressed striae, known as lands and grooves, on the bullet as it is fired from the weapon, whereas the striking of the firing pin against the base of the cartridge, which initiates discharge of the ammunition, will leave an impression but not striae.

...

An examiner observes three types of characteristics on spent bullets or cartridges: class, subclass and individual. Class characteristics are gross features common to most if not all bullets and cartridge cases fired from a type of firearm, for example, the caliber and the number of lands and grooves on a bullet. Individual characteristics are microscopic markings produced in the manufacturing process by the random imperfections of tool surfaces (the constantly changing tool as described above) and by use of and/or damage to the gun post-manufacture. . . . Subclass characteristics generally fill the gap between the class and individual characteristics categories. They are produced incidental to manufacture but apply only to a subset of the firearms produced, for example, as may occur when a batch of barrels is formed by the same irregular tool.

Id. at 427-28; see also Assoc. of Firearm and Tool Mark Examiners, Theory of Identification, Range of Striae Comparison Reports and Modified Glossary Definitions – An AFTE Criteria for Identification Committee Report, 24 AFTE Journal 336, 340 (1992) (“AFTE 1992 Theory”) (defining class, subclass, and individual characteristics).

Toolmark analysis is a “forensic feature-comparison method,” which is “a procedure by which an examiner seeks to determine whether an evidentiary sample (e.g., from a crime scene) is or is not associated with a source sample (e.g., from a suspect) based on similar features.”² President’s Council of Advisors on Sci. & Tech., Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods, 46 (2016) (“PCAST Report”) (available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf) (last visited Nov. 21, 2019); see also AFTE 1992 Theory, 24 AFTE Journal at 339 (defining “toolmark identification” as “a discipline of forensic science”). Toolmark analysis, along with other feature-comparison methods, belongs to the “scientific discipline [of] metrology, which is ‘the science of measurement and its application;’” here, firearms examiners measure and compare impressions, striae, and other toolmarks on different pieces of ballistics evidence. PCAST Report at 23 (quoting Joint Committee for Guides in Metrology, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, JCGM 200, 16 (3rd ed. 2012) (available at https://www.bipm.org/utis/common/documents/jcgm/JCGM_200_2012.pdf)) (emphasis omitted) (last visited Nov. 21, 2019); cf. id. at 44 n.93 (“That forensic feature-comparison methods belong to the field of metrology is clear from the fact that the [National Institute of Standards and Technology] . . . ,

² Other “feature-comparison” methods include comparative analysis of bite marks, fingerprints, and DNA. PCAST Report at 46.

which is the world's leading metrological laboratory[,] is the home within the Federal government for research efforts on forensics science.”).

2. The AFTE and the “Sufficient Agreement” Standard

The Association of Firearms and Toolmark Examiners (the “AFTE”) is the “international professional organization for practitioners of [f]irearm and/or [t]oolmark [i]dentification.” What is AFTE?, <http://afte.org/about-us/what-is-afte> (last visited November 21, 2019). “Membership in the [AFTE] is limited to those persons of integrity with suitable education, training, and experience in the examination of firearms and/or toolmarks.” Membership Requirements, <http://afte.org/membership/membership-requirements> (last visited November 21, 2019).

The “AFTE Theory” is “a theory of toolmark identification adopted by the [AFTE].” United States v. Sebborn, No. 10-CR-87 (SLT), 2012 WL 5989813, at *3 (E.D.N.Y. Nov. 30, 2012). Under the AFTE Theory, an examiner comparing two pieces of ballistics evidence may reach one of four conclusions: (1) “identification,” meaning the pieces of evidence come from the same source; (2) “elimination,” meaning that they came from different sources; (3) “inconclusive,” meaning that there is not enough evidence for an examiner to make a determination; and (4) “unsuitable,” which means that the recovered evidence lacks discernable class and individual characteristics. See AFTE 1992 Theory, 24 AFTE Journal at 337-38.

The AFTE standard for an “identification” determination is “sufficient agreement” between two pieces of evidence. The AFTE defines sufficient agreement as follows:

“[S]ufficient agreement” is related to the significant duplication of random toolmarks as evidenced by the correspondence of a pattern or combination of patterns of surface contours. Significance is determined by the comparative examination of two or more sets of surface contour patterns comprised of individual peaks, ridges and furrows. Specifically, the relative height or depth, width, curvature and spatial relationship of the individual peaks, ridges and furrows within one set of surface contours are defined and compared to the

corresponding features in the second set of surface contours. Agreement is significant when the agreement in individual characteristics exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool. The statement that “sufficient agreement” exists between two toolmarks means that the agreement of individual characteristics is of a quantity and quality that the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility.

Assoc. of Firearm and Tool Mark Examiners, Theory of Identification as it Relates to Tool Marks: Revised, 43 AFTE Journal 4, 287 (2011) (“AFTE Revised Theory of Identification”)

(emphasis omitted). The AFTE Theory explains that the act of determining whether two pieces of ballistics evidence came from the same source is “[c]urrently . . . subjective in nature, founded on scientific principles and based on the examiner’s training and experience.” Id.

D. Scientific Discussion Around Toolmark Analysis

1. Recent Scientific Publications Discussing Toolmark Analysis

Firearm toolmark analysis has been scrutinized by at least two scientific reports published in the past ten years. See PCAST Report; Comm. on Identifying the Needs of the Forensic Scis. Cmty., Nat’l Research Council, Strengthening Forensic Science in the United States: A Path Forward (2009) (“NRC Report”) (available at <https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf>) (last visited Nov. 21, 2019).

a. The NRC Report

The NRC Report was published in 2009 with an intent to “chart an agenda for progress in the forensic science community and its scientific disciplines.” Id. at xix. It considered several branches of forensic science, including firearms toolmark analysis. See, e.g., id. at 3-4. After reviewing the theory underlying toolmark analysis, id. at 150-51, the NRC Report noted several weaknesses in the field: “Knowing the extent of agreement in marks made by different tools, and

the extent of variation in marks made by the same tool, is a challenging task.” Id. at 153. The report noted the potential for new technology or techniques to improve accuracy but emphasized that “the decision of the toolmark examiner remains a subjective decision based on unarticulated standards and no statistical foundation for the estimation of error rates.” Id. at 153-54. The report concluded that “[i]ndividual patterns from manufacture or wear might, in some cases, be distinctive enough to suggest one particular source, but additional studies should be performed to make the process of individualization more precise and repeatable.” Id. at 154.

b. The PCAST Report

The PCAST Report was published in 2016, seven years after the NRC Report. The report aimed to provide “clarity about the scientific standards for the validity and reliability of forensic methods” and to “evaluate specific forensic methods to determine whether they have been scientifically established to be valid and reliable.” PCAST Report at 1. The PCAST Report analyzed the scientific research on seven branches of forensic sciences, including firearms analysis. Id. at 7-14. The report sought to establish whether each branch of forensic science had achieved (1) “foundational validity,” which the report defined as “the scientific standard corresponding to the legal standard of evidence being based on ‘reliable principles and methods’” and (2) “validity as applied,” which the report defined as “the scientific standard corresponding to the legal standard of an expert having ‘reliably applied the principles and methods.’” Id. at 43 (quoting Fed. R. Evid. 702) (emphasis omitted). The report identified “two key elements” of foundational validity. Id. at 48. First, a method must have “a reproducible and consistent procedure for (a) identifying features within evidence samples; (b) comparing the features in two samples; and (c) determining . . . whether the samples should be declared to be a [match].” Id. Second, there must be “empirical measurements, from multiple independent

studies, of (a) the method's false positive rate . . . [and] (b) the method's sensitivity," which is the "probability that it declares a [match] between samples that actually come from the same source." Id. (emphasis omitted).

Finally, the report noted a difference between "objective" and "subjective" forensic science methods: objective methods "consist[] of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising little or no judgment," and subjective methods "includ[e] key procedures that involve significant human judgment." Id. at 47. For subjective methods like firearms toolmark analysis, see id. at 104, "the foundational validity . . . can be established only through empirical studies of examiner's performance to determine whether they can provide accurate answers" because "the black box in the examiner's head cannot be examined directly for its foundational basis in science." Id. at 49 (emphasis in original).

The report found that "firearms analysis currently falls short of the criteria for foundational validity." Id. at 112. The report asserted that the "sufficient agreement" standard falls short of the necessary "reproducible and consistent procedure" to reliably compare forensics evidence; the report characterized the AFTE Theory of Identification ("AFTE Theory") as "clearly not a scientific theory" and "circular." Id. at 60. The report also found that the toolmark analysis failed to meet the second key element of foundational validity. It analyzed several studies with different study designs, id. at 106-10, and concluded that there was "only a single study that was appropriately designed to test foundational validity and estimate reliability," id. at 111. It concluded by observing the "need for additional, appropriately designed . . . studies to provide estimates of reliability." Id.

2. Foundational Scientific Validity and Evidentiary Reliability Under Daubert

The PCAST Report determined that toolmark analysis lacks foundational scientific validity, and Mr. Shipp relies on this to argue that Detective Ring's testimony should be excluded in its entirety. (Mem. at 10-12.) The government, in turn, dismisses the PCAST Report as adding nothing new to the analysis of the prior judicial decisions that have found toolmark analysis to be reliable. (Gov't Opp'n at 7, 10-11).

Therefore, the court must initially decide whether the AFTE Theory lacking foundational scientific validity, if true, would necessarily mean that Detective Ring is unable to provide expert testimony. Courts have generally answered this question "no." Some decisions unambiguously disclaim the importance of whether toolmark analysis is scientifically valid. See, e.g., Otero, 849 F. Supp. 2d. at 430-31 ("This Court expresses no opinion on whether the practice of firearms and toolmark identification constitutes a 'scientific' discipline because that is not the question before the court."); see also United States v. Johnson, No. 16-CR-281 (PGG), 2019 WL 1130258, at *5 (S.D.N.Y. March 11, 2019) (citing Otero with approval). Several other courts have, without explicitly connecting the two issues, found that the AFTE Theory was not scientifically valid but was reliable. See, e.g., United States v. Simmons, No. 16-CR-130, 2018 WL 1882827, at *4, *9 (E.D. Va. Jan. 12, 2018), R&R adopted by, 2018 WL 658693 (E.D. Va. Feb. 1, 2018); Ashburn, 88 F. Supp. 3d. at 248-49; United States v. Taylor, 663 F. Supp 2d. 1170, 1179-80 (D.N.M. 2009); Glynn, 578 F. Supp. 2d at 570-71.³

³ Most of these decisions were rendered before one or both of the NRC (2009) and PCAST (2016) reports were published.

In reviewing this case law, the court notices a tension between Daubert and courts' tendency to separate the AFTE Theory's scientific validity and the evidentiary reliability of toolmark analysis. Daubert emphasized that the "overarching subject" of the Rule 702 inquiry "is the scientific validity—and thus the evidentiary reliability—of the principles that underline a proposed submission." Daubert, 509 U.S. at 594-95. Kumho Tire, in applying Daubert to nonscientific expert testimony, observed that Rule 702 "makes no relevant distinction between 'scientific' knowledge and 'technical' or 'other specialized' knowledge . . . [and] makes clear that any such knowledge might become the subject of expert testimony." 526 U.S. at 147. Thus, while Kumho Tire broadened the scope of the Daubert inquiry, it did not alter Daubert's standard of reliability for scientific testimony. See id. at 147-48 (noting that Daubert "referred to scientific testimony because that was the nature of the expertise at issue" (quotation marks omitted) (alterations adopted)); see also Daubert, 509 U.S. at 590 n.9 ("In a case involving scientific evidence, evidentiary reliability will be based upon scientific validity." (emphasis in original)).

Recently, however, the Second Circuit affirmed a district court's decision to admit expert testimony about a certain type of hair analysis even though certain aspects of the technique "had not been established to a degree of scientific certainty." Restivo, 846 F.3d at 576. The Second Circuit explained that "[t]here is no basis in [Rule 702] or in . . . case law to suggest that a scientist whose testimony could not pass muster under Daubert as 'scientific knowledge' could not testify to their technical or other specialized knowledge, so long as that testimony was reliable." Id. (emphasis added).

Here, Defendant seeks to exclude Detective Ring's testimony in its entirety, arguing, in substance, that Detective Ring is offering scientific evidence based on a scientifically invalid

technique. (Mem. at 10-12.) The thrust of this argument is misplaced. Even accepting that conclusions based on the AFTE Theory are “scientific evidence” and that the AFTE Theory lacks foundational scientific validity, Defendant does not necessarily prevail. In that case, Detective Ring could still testify based on his “specialized knowledge” as a firearms toolmark examiner as long as the court determines that his testimony is reliable. Restivo, 846 F.3d at 576.

Still, the court may consider critiques of a method’s validity when determining the method’s reliability. See id. at 575 (“[T]he district court has broad discretion in determining what method is appropriate for evaluating reliability under the circumstances of each case.” (quoting Amorgianos, 303 F.3d at 265)). A method or technique that is scientifically valid will almost certainly also be reliable. See Daubert, 509 U.S. at 590 n.9. Therefore, the court finds that the three characteristics of foundational scientific validity identified by the PCAST Report—repeatability, reproducibility, and accuracy⁴—are instructive for assessing the reliability of toolmark analysis. See PCAST Report at 47. People may hold different views on whether the AFTE Theory is scientifically valid. However, it is uncontroversial that toolmark analysis testimony should not be admitted if, for example, examiners reach different conclusions when examining different evidence from the same firearm⁵ (i.e., the conclusions must be repeatable); different examiners reach different conclusions (i.e., the conclusions must be reproducible); or examiners make incorrect conclusions (i.e., the conclusions must be accurate).

⁴ A method is repeatable if “an examiner obtains the same result[] when analyzing samples from the same sources.” PCAST Report at 47. It is reproducible if “different examiners obtain the same result[] when analyzing the same samples.” Id. It is accurate if “an examiner obtains correct results both (1) for samples from the same source (true positives) and (2) for samples from different sources (true negatives).” Id.

⁵ To clarify, the concern here is not that an examiner might find some pieces of ballistics evidence unsuitable for examination or inconclusive, but that an examiner might make contradictory conclusions of “identification” and “elimination” when comparing two pieces of ballistics evidence from the same firearm.

Therefore, while the purported invalidity of the AFTE Theory does not preclude Detective Ring from testifying, the court seriously considers the PCAST Report's critiques when assessing the reliability of Detective Ring's proposed testimony.

E. Application of the Daubert Factors to Toolmark Analysis

The Government's argument for the admission of Detective Ring's testimony rests largely on the fact that other courts have considered and admitted expert testimony on toolmark analysis. (See Gov't Opp'n at 4 ("The [D]efendant's initial argument . . . has been rejected by every known court to consider it . . .").) Therefore, the court will review this case law before discussing how each of the Daubert factors applies to the AFTE Theory.

The Government is correct that several courts have considered whether toolmark analysis is reliable and whether toolmark analysis testimony is admissible under Federal Rule of Evidence 702. See Johnson, 2019 WL 1130258, at *12-13 (collecting cases). The Government is also correct that, without exception, these courts have admitted the expert testimony. Id.; (see also Mem. at 12; Gov't Opp'n at 14). However, many of these courts have expressed increasing concerns about the scientific validity of toolmark analysis, and have imposed restrictions on the proffered expert testimony. See, e.g., United States v. White, No. 17-CR-611 (RWS), 2018 WL 4565140, at *3 (S.D.N.Y. Sept. 24, 2018); United States v. Simmons, 2018 WL 1882827, at *8; Ashburn, 88 F. Supp. 3d at 249; Taylor, 663 F. Supp. 2d at 1180; Glynn, 578 F. Supp 2d at 574-75.

Several of these decisions were issued prior to the publication of the NRC Report in 2009. See Johnson, 2019 WL 1130258, at *12-13 (listing cases). Very few were issued after the publication of the PCAST Report in 2016. See United States v. Romero-Lobato, 379 F. Supp. 3d 1111 (D. Nev. 2019); Johnson, 2019 WL 1130258; United States v. Hylton, No. 17-CR-00086,

2018 WL 5795799 (D. Nev. Nov. 5, 2018); White, 2018 WL 4565140; Simmons, 2018 WL 1882827. Of these, only Romero-Lobato and Johnson discussed the PCAST Report, and both did so summarily. Romero-Lobato, 379 F. Supp. 3d at 1117-18; Johnson, 2019 WL 1130258, at *11, *15.

The court's "discretion in choosing the manner of testing expert reliability . . . is not discretion to abandon the gatekeeping function . . . [or] perform the function inadequately." Kumho Tire, 526 U.S. at 158-59 (Scalia, J. concurring). Even though prior decisions have found toolmark analysis to be reliable, it is incumbent upon this court to thoroughly review the critiques of the AFTE Theory found in the NRC and PCAST Reports and to consider whether they merit exclusion of Detective Ring's testimony or, alternatively, appropriate limitations on his testimony.

1. Whether the AFTE Theory Can Be and Has Been Tested

The first factor under Daubert is whether a technique "can be (and has been) tested." Daubert, 509 U.S. at 592. As the Government explains (see Gov't Opp'n at 9), and several courts have found, the AFTE Theory has been subjected to considerable testing, especially since Daubert and Kumho Tire. See Romero-Lobato, 379 F. Supp 3d. at 1118-19; Ashburn, 88 F. Supp. 3d at 245; Otero, 849 F. Supp. 2d at 432-33. Both the AFTE Website and the PCAST Report discuss some of the relevant studies. See Testability of the Scientific Principle, AFTE, <http://afte.org/resources/swggun-ark/testability-of-the-scientific-principle> (last visited Nov. 21, 2019); PCAST Report at 106-12.

Defendant is correct to point out that not all of the studies are equally probative of the AFTE Theory's reliability (Mem. at 10), and courts should avoid placing too much trust in studies that may overestimate the method's reliability. However, the probative value of different

study designs is more appropriately considered as part of the discussion of the method's error rate, below. The court finds that the AFTE Theory can be and has been tested and this factor therefore weighs in favor of reliability.

2. Whether Toolmark Analysis Has Been Subjected to Peer Review

The next Daubert factor is whether the AFTE Theory has been subjected to "peer review and publication." Daubert, 509 U.S. at 594. Prior decisions have, with near uniformity, determined that the AFTE Theory has been subjected to peer review and found this factor to weigh in favor of a finding of admissibility. See, e.g., Johnson, 2019 WL 113025, at *16; Ashburn, 88 F. Supp. 3d at 246; Otero, 849 F. Supp. 2d at 433; United States v. Monteiro, 407 F. Supp. 2d 351, 367 (D. Mass. 2006). Many of these decisions place substantial weight on the AFTE Journal and the articles published therein. See, e.g., Johnson, 2019 WL 113025, at *16 (reviewing case law and finding that "[c]ourts addressing this Daubert factor have determined that the AFTE Journal scholarship qualifies as peer-reviewed literature").

However, United States v. Tibbs, No. 2016-CF1-19431, 2019 WL 4359486 (D.C. Super. Sep. 5, 2019), recently challenged the quality of the AFTE Journal's peer review process. This thorough opinion includes several pages analyzing the AFTE Journal's peer review process and highlights several reasons for assigning less weight to articles published in the AFTE Journal than in other publications.⁶ Id. at *8-*10.

⁶ Tibbs took issue with three aspects of the AFTE Journal peer review process: (1) The AFTE Journal employs an "open" instead of "double-blind" peer review process, *i.e.*, the review process is not anonymous; (2) the "AFTE does not make [its journal] generally available to the public or the world of possible reviewers and commentators outside of the organization's membership;" and (3) articles proposed for publication are reviewed by members of the editorial board "composed entirely of members of AFTE," who "may be trained and experienced in the field of firearms and toolmark examination, but do not necessarily have any specialized or even relevant training in research design and methodology" and who "have a vested interest in publishing studies that validate their own field and methodologies." Tibbs, 2019 WL 4359486, at *9-*10.

The court shares these concerns about the AFTE Journal's peer review process. In particular, the court is concerned that the reviewers, who are all members of the AFTE, "have a vested, career-based interest in publishing studies that validate their own field and methodologies." Id. at *10. Also concerning is the possibility that the reviewers "may be trained and experienced in the field of firearms and toolmark identification, but [may] not necessarily have any specialized or even relevant training in research design and methodology." Id.

However, even assigning limited weight to the substantial fraction of the literature that is published in the AFTE Journal, this factor still weighs in favor of admissibility. Daubert found the existence of peer-reviewed literature important because "submission to the scrutiny of the scientific community . . . increases the likelihood that substantive flaws in the methodology will be detected." Daubert, 509 U.S. at 593. Despite AFTE Journal's open peer-review process, the AFTE Theory has still been subjected to significant scrutiny. Indeed, the scrutiny of PCAST and the flaws it perceived in the AFTE Theory inform much of the discussion of the next two Daubert factors. Therefore, the court finds that the AFTE Theory has been sufficiently subjected to "peer review and publication." Daubert, 509 U.S. at 594.

3. The Error Rate for Toolmark Analysis

The court next considers the "known or potential rate of error" of the AFTE Theory. Daubert, 509 U.S. at 594. Defendant attempts to limit discussion of this prong by arguing that there has only been "one black box study" of firearms toolmark analysis and that "a single study estimating [the] error rate is insufficient to determine a known rate of error for the field." (Mem. at 11.) However, Daubert instructs the court to consider potential rates of error as well as

known error rates, 509 U.S. at 594, and therefore a broader discussion of estimated error rates is merited.

The Government acknowledges that it is “difficult to establish” a known error rate for toolmark analysis (Gov’t Opp’n at 11) but rests on judicial decisions finding that “[s]tudies have shown that the error rate among trained toolmark and firearm examiners is quite low.” Ashburn, 88 F. Supp. 3d at 246; see also Romero-Lobato, 379 F. Supp. 3d at 1119-20; Johnson, 2019 WL 1130258, at *18-*19. The studies referenced by these decisions found the potential error rate was roughly between 1%-2%.⁷ See e.g., Romero-Lobato, 379 F. Supp. 3d at 1119-20; Johnson, 2019 WL 1130258, at *18-*19.

The PCAST Report examined several of these studies and cautioned that not all study designs are equally trustworthy when predicting error rates. PCAST Report at 106-11. Many of the studies employed a “closed-set” design, wherein examiners are given two sets of bullets or shell casings with matching bullets or casings split between the two sets. Id. at 108. The examiners are then asked to perform a matching exercise between the two sets. Id. This design is “simpler than the problem encountered in casework, because the correct answer is always present in the collection.” Id. “[E]xaminers can perform perfectly if they simply match each bullet to the standard that is closest.” Id. (emphasis in original). Additionally, each identified match limits the options for the remaining bullets or casings, which the Director of the Defense Forensic Science Center likened to “solving a Sudoku puzzle where initial answers can be used to help fill in subsequent answers.” Id. at 106. With all of these advantages—none of which are

⁷ Additionally, one court considered testimony that a European study “that mimic[ed] the imperfect samples often found in the field yielded an error rate of around 5%,” but found that error rate was “not excessively high.” United States v. McCluskey, No. CR 10-27340, 2013 WL 12335325, at *7 (D.N.M. Feb. 7, 2013). This decision is not determinative of the court’s holding here, but the court is uncomfortable with the finding that an error rate that equates to one false positive out of every twenty examinations weighs in favor of admission. Id.

present in fieldwork—it is unsurprising that studies with this design found a false positive rate of less than one tenth of one percent. Id. at 111.

PCAST next considered one study with a “partly-open set” design, which was similar to the “closed set” design, except that two of the 15 samples came from a firearm “for which known standards were not provided.”⁸ Id. at 109. The false-positive rate for these two samples was 2.1%, or “roughly 100-fold higher” than the false-positive rate from the “closed set” studies. Id.

Finally, PCAST analyzed an independent-set or “black-box” study.⁹ Id. at 110. In this study, examiners were “presented with 15 separate comparison problems—each consisting of one questioned sample and three known test fires from the same known gun, which might or might not have been the source.” Id. (emphasis in original). This study therefore required examiners to make each determination based only on the comparison of the questioned sample to its three test fires. The independence of each comparison means that an identification determination on one problem has no effect on the analysis of the remaining problems. Id. This design is the most similar to the situation examiners face in fieldwork. This study estimated an error rate as high as 2.2%, or one in every 46 comparisons. Id. at 110-11.¹⁰

⁸ Thomas G. Fadul Jr., Gabriel A. Hernandez, Erin Wilson, Stephanie Stoiloff & Sneha Gulati, An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing Consecutively Manufactured Glock EBIS Barrels with the Same EBIS Pattern, National Institute of Justice (2013) (available at <https://www.ncjrs.gov/pdffiles1/nij/grants/244232.pdf>) (last visited Nov. 21, 2019).

⁹ David P. Baldwin, Stanley J. Bajic, Max Morris & Daniel Zamzow, A Study of False-Positive and False-Negative Error Rates in Cartridge Case Comparisons, Ames Laboratory, USDOE (2014) (available at <https://afte.org/uploads/documents/swggun-false-positive-false-negative-usdoe.pdf>) (last visited Nov. 21, 2019). This is the only study the PCAST Report found to have an “appropriate[]” design for determining reliability of toolmark analysis. PCAST Report at 111.

¹⁰ PCAST focused on the “false positive rates based on the proportion of conclusive examinations.” PCAST Report at 106. This is reasonable, as “evidence used against a defendant in court will typically be the result of a conclusive examination.” Id. at 91-92, 106. However, it is also worth noting that the rate of “inconclusive” determinations skyrocketed in the “partly-open” and “black box” studies. Id. at 112 (noting 41.8% and 33.7% inconclusive rates for the “partly open” and “black box” studies, respectively, as compared to .2% percent for the “closed-set” studies). While it is desirable for examiners to make an inconclusive determination when they feel it is merited, the vast difference in inconclusive rates casts further doubt on the miniscule error rates reported by the “close-set” studies.

Therefore, the study that most closely resembles fieldwork estimated that a firearms toolmark examiner may incorrectly conclude that a recovered piece of ballistics evidence matches a test fire once out of every 46 examinations. When compared to the error rates of other branches of forensic science—as rare as 1 in 10 billion for single source or simple mixture DNA comparisons, see id. at 72-73—this error rate cautions against the reliability of the AFTE Theory.¹¹

Based on the above information, the court finds that the potential rate of error for matching ballistics evidence based on the AFTE Theory does not favor a finding of reliability at this time. The court notes, however, that the FBI and the Ames Laboratory are currently conducting a second black box study on the AFTE Theory. See FBI-Ames Laboratory Blackbox Study FAQs, <https://www.ameslab.gov/operations/faq/fbi-ames-laboratory-blackbox-study-faqs> (last visited November 21, 2019). The results of this and future studies may, of course, change much of the foregoing analysis.

4. The AFTE Theory as a Controlling Standard

The court next considers the “existence and maintenance of standards controlling the technique’s operation.” Daubert, 509 U.S. at 594. Here, the court considers the AFTE Theory of Identification. See Sebborn, 2012 WL 5989813, at *3. As described above, an examiner applying the AFTE Theory may conclude two pieces of evidence to be of “common origin . . .

¹¹ Defendant argues that the court should also consider the error rates for the NYPD laboratory and for Detective Ring. (Mem. at 11, 14.) Consideration of these individualized error rates is more pertinent for the determining whether an “expert reliably applied the principles and methods to the facts of the case.” Fed. R. Evid. 702(d). In response to a court order (see Oct. 3, 2019 Min. Entry), the Government provided reports showing that Detective Ring achieved “successful completion” on all of his proficiency tests. (See NYPD Police Lab. Performance Monitoring and Proficiency Test Results for Detectives Ring (Dkt. 36-2).) However, before the court could make a determination that Detective Ring reliably applied the method, it would need to know more about the structure of the exam and whether Detective Ring made any false-positive identifications. The limitations the court places on Detective Ring’s testimony mitigates the need for additional information on the structure of the NYPD laboratory firearms analysis proficiency tests.

when the unique surface contours of two toolmarks are in ‘sufficient agreement.’” AFTE Revised Theory of Identification, 43 AFTE Journal at 287.

“[B]oth courts and the scientific community have voiced serious concerns about the ‘sufficient agreement’ standard, characterizing it as ‘tautological,’ ‘wholly subjective,’ ‘circular,’ ‘leaving much to be desired,’ and ‘not scientific.’” Johnson, 2019 WL 113025, at *17 (alteration adopted); see also William A. Tobin & Peter J. Blau, Hypothesis Testing of the Critical Underlying Premise of Discernible Uniqueness in Firearms-Toolmarks Forensic Practice, 53 Jurimetrics J. 121, 127 (2013) (“Because the [AFTE Theory] is comprised of such vague and subjective terms, with no underlying protocol, it does not incorporate, or even allow for, two critical cornerstones of true scientific endeavor: repeatability and reproducibility.”). In its opposition, the Government notes that “courts have acknowledged that this is the Daubert factor on which firearm toolmark analysis scores the lowest.” (Gov’t Opp’n at 12.) The criticisms of the AFTE Theory appear well founded, and the court notes two fundamental issues.

First, the sufficient agreement standard is circular and subjective. Reduced to its simplest terms, the AFTE Theory “declares that an examiner may state that two toolmarks have a ‘common origin’ when their features are in ‘sufficient agreement.’” PCAST Report at 60. “It then defines ‘sufficient agreement’ as occurring when the examiner considers it a ‘practical impossibility’ that the toolmarks have different origins.” Id. The NRC Report notes that the AFTE Theory “is the best guidance available for the field of toolmark identification, [but] does not even consider, let alone address, questions regarding variability, reliability, repeatability, or the number of correlations needed to achieve a given degree of confidence.” NRC Report at 155. Without guidance as to the extent of commonality necessary to find “sufficient agreement,” the AFTE Theory instructs examiners to draw identification conclusions from what is essentially a

hunch—a hunch “based on the examiner’s training and experience,” AFTE Revised Theory of Identification, 43 AFTE Journal at 287—but still a hunch.

Moreover, the application of this circular standard is “subjective in nature . . . based on the examiner’s training and experience.” AFTE Revised Theory of Identification, 43 AFTE Journal at 287. Ostensibly, one hundred firearms toolmark examiners could hold one hundred different personal standards of when two sets of toolmarks sufficiently agree, and all one hundred of these personal standards may accord with the AFTE Theory. Further, because the standard itself offers so little guidance on when an examiner should make an identification determination, some examiners may decide that the two sets of toolmarks were made by the same tool while others determine the toolmarks to be inconclusive and still others decide the toolmarks were made by different tools. To emphasize, these one hundred examiners could come to these contradictory conclusions without a single examiner running afoul of the AFTE Theory.¹²

Some decisions minimize the import of the subjective nature of the standard. See, e.g., Simmons, 2018 WL 1882827 at *5 (“[A]ll technical fields which require the testimony of expert witnesses engender some degree of subjectivity requiring the expert to employ his or her individual judgment, which is based on specialized training, education, and relevant work experience.” (emphasis in original)); see also Romero-Lobato, 379 F. Supp 3d at 1120 (noting that excluding opinions “derived from subjective methodology . . . would, in most circumstances, exclude psychologists, physicians, and lawyers from testifying as expert witnesses.”).

¹² The risk is lessened where, as here, a second examiner independently confirms the results. However, because the second examiner applies the same subjective standard without any additional guidance, the confirmation does not eliminate the inherent risk associated with the application of a circular and subjective controlling standard.

There is a difference, though, between “some degree of subjectivity,” as exists when a medical expert testifies as to whether a doctor met a certain accepted standard of care, and the near total subjectivity countenanced by the AFTE Theory, where there is no actual guidance for what comprises “sufficient agreement.” Moreover, unlike psychologists, physicians, and lawyers, toolmark examiners have little opportunity to apply the AFTE Theory outside of judicial proceedings and criminal investigations. It is difficult, therefore, to establish whether an examiner “employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field.” Kumho Tire, 526 U.S. at 152. Finally, there is a distinction between, for example, a psychologist testifying as to whether a defendant is competent to stand trial and a firearms examiner testifying as to whether two bullets were fired from the same firearm. In the former, the expert’s testimony sheds light on an inherently ambiguous question about which qualified experts may reasonably disagree. In the latter, the expert is answering an unambiguous question: Do two pieces of ballistics evidence share the same source firearm or do they not?¹³

Second, and relatedly, the court has serious concerns about the ability of firearms toolmark examiners to protect against false positives due to, for example, random similarities in two sets of toolmarks or an accidental misidentification of subclass characteristics as individual

¹³ In an attempt to reduce the subjectivity of the analysis, some examiners and labs have adopted the consecutive matching striae (“CMS”) technique. “Under CMS, the threshold for identifying a particular tool as the source of a three-dimensional toolmark is a match between evidence and test toolmarks of one group of six consecutive matching striae or two different groups of at least three consecutive matching striae in the same relative position.” See Adina Schwartz, A Systemic Challenge to the Reliability and Admissibility of Firearms and Toolmark Identification, 6 Colum. Sci. & Tech. L. Rev. 2, 39 (2005).; see also United States v. Johnson, 2019 WL 1130258, at *9, *18. While the CMS technique lessens the inherent subjectivity of the AFTE Theory, it depends on the examiner matching only individual characteristics (as opposes to subclass or class characteristics); the identification of similar striae and the characterization of a certain mark as an individual or subclass characteristic remains a subjective decision. See Schwartz, 6 Colum. Sci. & Tech. L. Rev. at 40. Additionally, “the necessary research has not been done” to determine the reliability of the CMS technique. *Id.* at 41. Therefore, the CMS technique is indicative of the potential to develop an objective, uniform standard for examiners to make identification determinations, but it has not been incorporated into the AFTE Theory and its repeatability, reproducibility and accuracy have yet to be established.

characteristics. See Monteiro, 407 F. Supp. 2d at 371 (“[O]ne critical problem with the AFTE Theory is the lack of objective standards for deciding whether a particular mark is a subclass or individual characteristic.”). The AFTE Theory instructs examiners to conclude that markings sufficiently agree only when the similarity in two sets of markings “exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools.” AFTE Revised Theory of Identification, 43 AFTE Journal at 287. The theory, however, provides no additional guidance as to what “the best agreement demonstrated between toolmarks known to have been produced by different tools” actually is. Id. The AFTE Theory thus provides its practitioners with a theoretical standard for the minimum agreement between two pieces of evidence, a level of agreement that—if absent—precludes an examiner from making an identification determination. However, there is no practical guidance for what this minimum level of agreement looks like in practice because the closest possible resemblance between different-sourced pieces of ballistics evidence is unknown. Therefore, there is limited support for the assertion it is close to impossible for two random firearms to produce toolmarks that “sufficiently agree” with each other.¹⁴ Id.

The determination that the similarity between two sets of toolmarks indicates sufficient agreement between them and is not, instead, a result of subclass characteristics or random similarities between different firearms is left to the examiner’s “training and experience.” AFTE Revised Theory of Identification, 43 AFTE Journal at 287; see also United States v. Monteiro,

¹⁴ Additionally, “relevant and representative databases” of the toolmarks left by various firearms “have not been developed.” Schwartz, 6 Colum. Sci. & Tech. L. Rev. at 46. Schwartz explains, in a discussion about the CMS technique, how developing large databases of the marks made by different firearms would help to realize CMS’s goal of “curing the absence of systemic knowledge of the differences and similarities between both (1) toolmarks produced by different tools of the same type and (2) toolmarks produced by the same tool.” Id. at 49. In the past, large representative databases helped scientists to develop reliable standards for matching DNA. Id. at 51. These databases are also helpful for proficiency testing; for example, the field of latent fingerprint analysis frequently relies on large searchable databases to test examiners with samples exhibiting the maximum similarity that can exist between non-matching fingerprints. See PCAST Report at 93-95.

No. 03-CR-10329, 2005 WL 8163021, at *5 (D. Mass. Nov. 28, 2005) (“[A]lthough the AFTE Theory indicates that ‘caution should be exercised in distinguishing subclass characteristics from individual characteristics,’ it does not offer any guidance for how an examiner should do that outside of relying on his or her individual training and experience.” (quoting AFTE 1992 Theory, 24 AFTE Journal at 340)). However, the PCAST Report argues against relying on an examiner’s training and experience when making a reliability determination:

“Experience” is an inadequate foundation for drawing judgments about whether two sets of features could have been produced by . . . different sources. Even if examiners could recall in sufficient detail all the patterns or sets of features that they have seen, they would have no way of knowing accurately in which cases two patterns actually came from different sources, because the correct answers are rarely known in casework.

...

“Training” is an even weaker foundation. The mere fact that an individual has been trained in a method does not mean that the method itself is scientifically valid nor that the individual is capable of producing reliable answers when applying the method.

PCAST Report at 61. This is a compelling argument. Detective Ring has “microscopically examined thousands of pieces of ballistics evidence,” but he examined them “for the purpose of aiding case detectives with criminal investigations and Assistant District Attorneys with criminal prosecutions.” (Ring CV.) This experience, while substantial, does not in itself grant Detective Ring reliable knowledge of “the best” agreement that may be generated by different firearms—even assuming that Detective Ring remembers the individual markings on each of the thousands of pieces of evidence that he has examined.

In sum, the subjectivity of the AFTE Theory raises serious concerns about the reproducibility of examination results across labs and examiners and the accuracy of those results. Therefore, the court finds that the subjective and circular nature of AFTE Theory weighs

against finding that a firearms examiner can reliably identify when two bullets or shell casings were fired from the same gun.

5. General Acceptance in the Relevant Scientific Community

Finally, the court must determine whether toolmark analysis has achieved general acceptance in the “relevant scientific community.” Daubert, 509 U.S. at 594. (“Widespread acceptance can be an important factor in ruling particular evidence admissible, and a known technique which has been able to attract only minimal support within the community may properly be viewed with skepticism.” (citations and quotation marks omitted)).

One important aspect of this factor is the definition of the “relevant scientific community.” Most courts have, in cursory fashion, identified toolmark examiners as the relevant community, and have summarily determined that the AFTE Theory is generally accepted in that community. See, e.g., Romero-Lobato, 379 F. Supp. 3d at 1122; Johnson, 2019 WL 1130258, at *19; Ashburn, 88 F. Supp. 3d at 247; Otero, 849 F. Supp. 2d at 435. Indeed, Romero-Lobato used this narrow definition of the relevant community to discount the criticism in the PCAST Report. See 379 F. Supp. 3d at 1122 (“[I]t is unclear if the PCAST Report would even constitute criticism from the relevant community because the committee behind the report did not include any members of the forensic ballistic community.” (quotation marks omitted)).

The court believes a broader definition is more appropriate for two main reasons. First, as mentioned above, firearms examiners “have a vested, career-based interest” in the AFTE Theory being accepted. Tibbs, 2019 WL 4359486, at *10. While the court in no way intends to equate toolmark analysis to the disciplines of “astrology or necromancy,” Kumho Tire, 526 U.S. at 151, the court takes guidance from Kumho Tire’s mandate to not assign an overly narrow definition to the relevant scientific community. See id. Second, as previously noted, the targeted

use of the AFTE Theory for criminal investigations and judicial proceedings limits the court's ability to assess whether an examiner "employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field." Kumho Tire, 526 U.S. at 152.

For these reasons, the court finds it appropriate to consider the opinions of the authors of the NRC Report and the PCAST Report who, while admittedly not members of the forensic ballistic community, are preeminent scientists and scholars and are undoubtedly capable of assessing the validity of a metrological method. See generally PCAST Report at v-ix; NRC Report at v-ix. As a result, the AFTE Theory has not achieved general acceptance in the relevant community, and this factor weighs against the reliability of the AFTE Theory.

6. Limitations on Detective Ring's testimony

On balance—because of the concerns raised by the lack of a known error rate with a potential error rate of one false positive per 46 examinations, the circular "sufficient agreement" standard, and, to a lesser extent, the lack of general acceptance in the scientific community—the Government has not demonstrated that the AFTE Theory of Identification is reliable enough to allow Detective Ring to testify that the recovered firearm is the source of the recovered bullet fragment and shell casing.

However, these concerns apply specifically and solely to use of the AFTE Theory to conclude that there is an identification, or match, between the test fires from the recovered firearm and the recovered shell casing and bullet fragment. They do not apply to several other aspects of Detective Ring's testimony. When examining the ballistics evidence that was recovered from the crime scene, Detective Ring took a number of steps, including, inter alia, (1) inspecting and testing the operability of the recovered firearm (Shipp 276), (2) inspecting the

recovered bullet fragments and shell casing and determining whether they were suitable for comparison with the test fires (Shipp 293, 298), (3) using a comparison microscope to compare a recovered bullet fragment and shell casing to test fires from the recovered firearm (Shipp 282-83, 291), and (4) identifying toolmarks on the recovered ballistics evidence that are similar to or consistent with the tool marks on the test fires. (Shipp 279, 289). Testimony about these and other related steps taken by Detective Ring is presumptively reliable and admissible under Rule 702 based on his specialized knowledge, training, and experience.

Therefore, the court will limit Detective Ring's testimony as follows: Detective Ring may testify as to his process of examining the recovered firearm, determining its operability, and test firing it. He may also describe the theory of toolmark analysis and how firearms can leave markings on bullets and shell casings. He may further describe the process of comparing the recovered shell casing and bullet fragments to the test fires and identify the similarities between them.¹⁵ Finally, he may testify that the toolmarks on the recovered bullet fragment and shell casing are consistent with having been fired from the recovered firearm, and that the recovered firearm cannot be excluded as the source of the recovered bullet fragment and shell casing. However, Detective Ring may not testify, to any degree of certainty, that the recovered firearm is the source of the recovered bullet fragment or the recovered shell casing.

This limitation is in line with, albeit slightly more restrictive than, limitations that other federal district courts have placed on toolmark analysis testimony.¹⁶ See, e.g., White, 2018 WL

¹⁵ And, of course, the defense may cross-examine Detective Ring about any inconsistencies between the recovered ballistics evidence and the test fires.

¹⁶ Additionally, this limitation is similar to the limitation placed by the D.C. Superior Court in Tibbs. See 2019 WL 4359486, at *23. Tibbs was issued after an evidentiary hearing that "involved detailed testimony from a number of distinguished expert witnesses, review of all of the leading studies in the discipline, [and] pre- and post-hearing briefing." Id. at *1. The 58-page opinion carefully reviewed the critiques and defenses of the AFTE Theory of Identification before concluding that "the government's [ballistics] expert may testify that[,] based on his

4565140, at *3 (precluding expert from testifying “to any specific degree of certainty as to his conclusion that there is a ballistics match”); Glynn, 578 F. Supp 2d at 574-75 (limiting expert’s testimony to stating that a match was “more likely than not”); see also Simmons, 2018 WL 1882827, at *8 (limiting testimony to a “a reasonable degree of ballistic . . . certainty”); Ashburn, 88 F. Supp. 3d at 249 (same); Taylor, 663 F. Supp. 2d at 1180 (same).

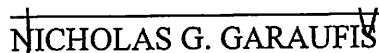
This more restrictive limitation is appropriate given the concerns raised by the PCAST Report about the lesser probative value of certain study designs and the reproducibility and accuracy of an individual examiner’s application of the “sufficient agreement” standard. Placing this limitation on Detective Ring’s testimony will prevent the jury from placing unwarranted faith in an identification conclusion based on the AFTE Theory, which the current research has yet to show can reliably determine, to a reasonable probability, whether separate pieces of ballistics evidence have the same source firearm. However, it will still allow the jury to benefit from Detective Ring’s extensive knowledge and experience examining ballistics evidence.

IV. CONCLUSION

For the foregoing reasons, Defendant’s (Dkt. 33) motion to exclude is DENIED and Detective Ring will be permitted to offer expert testimony in the area of firearms toolmark analysis subject to the limitations explained above.

SO ORDERED.

Dated: Brooklyn, New York
November 26, 2019

s/Nicholas G. Garaufis

NICHOLAS G. GARAUFIS
United States District Judge

examination, the recovered firearm cannot be excluded as the source of the cartridge casing found on the scene of the alleged shooting.” Id. at *23.